

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Artikel Penelitian

Systematic Literature Review: Optimasi Model Klasifikasi pada *Imbalanced Data* Menggunakan SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning*

Muhammad Alfin Ghozali ^a, Agus Bahtiar ^{b, *}^a Program Studi Teknik Informatika, STMIK IKMI Cirebon, Jl. Perjuangan No. 10 B, Karyamulya, Kesambi, Kota Cirebon, Jawa Barat 45135, Indonesia^b Program Studi Sistem Informasi, STMIK IKMI Cirebon, Jl. Perjuangan No. 10 B, Karyamulya, Kesambi, Kota Cirebon, Jawa Barat 45135, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 01 Juli 2026

Revisi Akhir: 20 Agustus 2026

Diterbitkan Online: 29 Agustus 2026

KATA KUNCI

Systematic Literature Review,
Imbalanced Data,
 SMOTE,
Hyperparameter Tuning,
Ensemble Learning

KORESPONDENSI

E-mail: muhammadalfin3008@gmail.com *

A B S T R A C T

Imbalanced Data merupakan salah satu permasalahan utama dalam pengembangan model klasifikasi karena menyebabkan model cenderung mempelajari kelas mayoritas sehingga kemampuan mendeteksi kelas minoritas menjadi menurun. Berbagai pendekatan, seperti *Synthetic Minority Oversampling Technique* (SMOTE), *Hyperparameter Tuning*, dan *Ensemble Learning*, telah dikembangkan untuk meningkatkan performa model pada kondisi tersebut. Penelitian ini bertujuan menganalisis karakteristik penelitian, mengkaji penerapan SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning*, serta mengidentifikasi *Research Gap* dalam optimasi model klasifikasi pada *Imbalanced Data*. Penelitian menggunakan metode *Systematic Literature Review* (SLR) dengan mengacu pada pedoman PRISMA 2020. Proses kajian meliputi penyusunan *Research Question*, pencarian literatur pada basis data Scopus, seleksi artikel berdasarkan kriteria inklusi dan eksklusi, *Quality Assessment*, ekstraksi data, serta sintesis hasil. Sebanyak 14 artikel yang memenuhi kriteria ditetapkan sebagai *Primary Studies*. Hasil sintesis menunjukkan bahwa SMOTE merupakan teknik yang paling banyak digunakan untuk menangani *Imbalanced Data*, sedangkan *Hyperparameter Tuning* berperan dalam memperoleh konfigurasi model yang optimal melalui pendekatan seperti *Optuna* dan *Bayesian Optimization*. Selain itu, *Ensemble Learning*, khususnya *Stacking* dan *Soft Voting*, banyak diintegrasikan dengan teknik optimasi lainnya untuk meningkatkan akurasi, stabilitas, dan kemampuan generalisasi model. Penelitian ini menyimpulkan bahwa integrasi SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning* menjadi pendekatan yang dominan dalam optimasi model klasifikasi pada *Imbalanced Data*. Penelitian juga mengidentifikasi peluang pengembangan pada aspek interpretabilitas model, efisiensi komputasi, serta pengembangan kerangka optimasi yang lebih terintegrasi.

1. PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi informasi mendorong penerapan *Machine Learning* pada berbagai bidang, seperti kesehatan, keuangan, lingkungan, pendidikan, keamanan siber, rekayasa perangkat lunak, hingga rekayasa material [1], [2]. Beragam algoritma klasifikasi dikembangkan untuk mendukung pengambilan keputusan melalui kemampuan mempelajari pola dari data dalam jumlah besar. Seiring meningkatnya kompleksitas data, pengembangan model klasifikasi tidak lagi hanya berfokus pada pemilihan algoritma, tetapi juga pada prapemrosesan data, optimasi parameter, dan evaluasi model untuk memperoleh akurasi, stabilitas, serta kemampuan generalisasi yang lebih baik [3], [4], [5], [6].

Salah satu tantangan utama dalam klasifikasi adalah *Imbalanced Data*, yaitu kondisi ketika distribusi data antar kelas tidak seimbang [7]. Pada kondisi ini, model cenderung mempelajari karakteristik kelas mayoritas sehingga kemampuan mendeteksi kelas minoritas menurun meskipun nilai *Accuracy* relatif tinggi [8]. Permasalahan tersebut banyak dijumpai pada diagnosis penyakit, deteksi penipuan, prediksi software fault, serta klasifikasi kualitas air dan lingkungan [9].

Berbagai pendekatan telah dikembangkan untuk mengatasi permasalahan tersebut. Salah satu metode yang paling banyak digunakan adalah *Synthetic Minority Oversampling Technique* (SMOTE), yaitu teknik *Oversampling* yang membangkitkan sampel sintetis pada kelas minoritas sehingga distribusi data menjadi lebih seimbang [10]. Dibandingkan *Random Oversampling*, SMOTE mampu meningkatkan representasi kelas

minoritas tanpa memperbesar risiko overfitting sehingga banyak diterapkan sebagai tahap prapemrosesan sebelum pelatihan model [11], [12], [13].

Selain penyeimbangan data, performa klasifikasi juga dipengaruhi oleh algoritma dan strategi optimasi yang digunakan. Berbagai penelitian menerapkan algoritma seperti *Decision Tree*, *Support Vector Machine (SVM)*, *K-Nearest Neighbor (KNN)*, *Naïve Bayes*, *Extreme Gradient Boosting (XGBoost)*, dan *Random Forest* [14]. Di sisi lain, pendekatan *Ensemble Learning*, seperti *Bagging*, *Boosting*, *Stacking*, dan *Soft Voting*, semakin banyak dimanfaatkan karena mampu meningkatkan akurasi dan kemampuan generalisasi model melalui kombinasi beberapa algoritma [5], [6], [15], [16].

Optimasi model juga memerlukan proses *Hyperparameter Tuning* karena penggunaan parameter bawaan belum tentu menghasilkan performa terbaik pada setiap dataset [3]. Berbagai metode, seperti *Grid Search*, *Random Search*, *Bayesian Optimization*, dan *Optuna*, telah digunakan untuk memperoleh konfigurasi parameter yang optimal. Beberapa penelitian bahkan mengombinasikan *Hyperparameter Tuning* dengan *Feature Selection* maupun *Cross Validation* untuk meningkatkan performa model secara lebih konsisten [13], [17], [18].

Meskipun penelitian mengenai optimasi model klasifikasi pada *Imbalanced Data* berkembang pesat [19], [20], pendekatan yang digunakan masih sangat beragam, baik dari sisi teknik penyeimbangan data, optimasi parameter, maupun penerapan *Ensemble Learning*. Perbedaan tersebut menyebabkan hasil penelitian sulit dibandingkan secara menyeluruh sehingga belum tersedia gambaran komprehensif mengenai kecenderungan metode yang digunakan maupun efektivitasnya pada berbagai domain aplikasi [16], [21].

Berdasarkan kondisi tersebut, penelitian ini menerapkan metode *Systematic Literature Review (SLR)* dengan pedoman PRISMA 2020 [22] untuk mengidentifikasi, mengevaluasi, dan mensintesis penelitian mengenai optimasi model klasifikasi pada *Imbalanced Data*. Kajian ini bertujuan menganalisis karakteristik penelitian, penerapan SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning*, sehingga dapat memberikan gambaran perkembangan penelitian sekaligus menjadi referensi bagi penelitian selanjutnya.

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, penelitian ini difokuskan pada empat *Research Question (RQ)*, yaitu:

1. Bagaimana karakteristik penelitian mengenai optimasi model klasifikasi pada *Imbalanced Data* menggunakan SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning*?
2. Bagaimana penerapan SMOTE dalam meningkatkan performa model klasifikasi pada *Imbalanced Data*?
3. Bagaimana metode *Hyperparameter Tuning* digunakan untuk mengoptimalkan performa model klasifikasi?
4. Bagaimana penerapan *Ensemble Learning* dalam meningkatkan performa model klasifikasi pada *Imbalanced Data*?

1.3. Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Mengidentifikasi karakteristik penelitian mengenai optimasi model klasifikasi pada *Imbalanced Data*.
2. Menganalisis penerapan SMOTE dalam meningkatkan performa model klasifikasi.
3. Mengidentifikasi metode *Hyperparameter Tuning* yang digunakan untuk mengoptimalkan performa model.
4. Menganalisis penerapan *Ensemble Learning* dalam meningkatkan akurasi dan kemampuan generalisasi model klasifikasi.

1.4. Kontribusi Penelitian

Penelitian ini memberikan sintesis mengenai perkembangan metode optimasi model klasifikasi pada *Imbalanced Data* melalui pendekatan SLR. Hasil kajian diharapkan menjadi referensi bagi peneliti dan praktisi dalam memahami kecenderungan penggunaan SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning*, serta menjadi dasar bagi pengembangan penelitian selanjutnya.

2. METODE

2.1. Metode Penelitian

Penelitian ini menggunakan metode *Systematic Literature Review (SLR)* untuk mengidentifikasi, mengevaluasi, dan mensintesis penelitian mengenai optimasi model klasifikasi pada *Imbalanced Data* melalui penerapan *Synthetic Minority Oversampling Technique (SMOTE)*, *Hyperparameter Tuning*, dan *Ensemble Learning*. Metode SLR dipilih karena mampu menghasilkan sintesis literatur secara sistematis, transparan, dan dapat direplikasi sehingga memberikan kesimpulan yang lebih objektif dibandingkan tinjauan pustaka konvensional [22].

Pelaksanaan penelitian mengacu pada pedoman PRISMA 2020 sebagai acuan dalam proses identifikasi, penyaringan, evaluasi kelayakan, hingga penetapan artikel yang digunakan sebagai *Primary Studies* [22]. Pendekatan ini mendukung proses seleksi literatur yang lebih transparan sekaligus meminimalkan potensi bias dalam pemilihan artikel.

Tahapan penelitian meliputi penyusunan *Research Question (RQ)*, strategi pencarian literatur, seleksi artikel berdasarkan kriteria inklusi dan eksklusi, *Quality Assessment (QA)*, ekstraksi data, serta sintesis hasil. Data yang diperoleh selanjutnya dianalisis secara deskriptif dan komparatif untuk mengidentifikasi karakteristik penelitian, penerapan metode optimasi, serta kecenderungan pengembangan model klasifikasi pada *Imbalanced Data*.

2.2. Tahapan Penelitian

Penelitian dilaksanakan melalui tahapan *Systematic Literature Review (SLR)* yang disusun secara sistematis sesuai pedoman PRISMA 2020 [22]. Proses diawali dengan identifikasi permasalahan dan penyusunan *Research Question* sebagai dasar

penentuan ruang lingkup penelitian. Selanjutnya disusun strategi pencarian literatur melalui penentuan basis data, search string, dan parameter pencarian yang sesuai dengan tujuan penelitian.

Artikel yang diperoleh kemudian diseleksi menggunakan kriteria inklusi dan eksklusi, dilanjutkan dengan *Quality Assessment* (QA) untuk memastikan kualitas metodologi setiap penelitian. Artikel yang memenuhi seluruh kriteria diekstraksi berdasarkan karakteristik penelitian, teknik penanganan *Imbalanced Data*, metode *Hyperparameter Tuning*, *Ensemble Learning*, serta metrik evaluasi yang digunakan. Data yang diperoleh selanjutnya dianalisis secara deskriptif dan komparatif untuk menjawab setiap *Research Question*.



Gambar 1. Diagram Tahapan Penelitian

Gambar 1 menunjukkan alur pelaksanaan *Systematic Literature Review* yang dimulai dari identifikasi permasalahan, penyusunan *Research Question*, pencarian literatur, seleksi artikel berdasarkan kriteria inklusi dan eksklusi, *Quality Assessment*, ekstraksi data, hingga sintesis hasil. Tahapan tersebut memastikan bahwa proses penelitian dilakukan secara sistematis, transparan, dan dapat direplikasi.

2.3. Research Question

Research Question (RQ) disusun sebagai dasar dalam pelaksanaan *Systematic Literature Review* (SLR). RQ berfungsi mengarahkan proses pencarian literatur, seleksi artikel, ekstraksi data, hingga sintesis hasil sehingga analisis yang dilakukan tetap sesuai dengan tujuan penelitian [22]. Pada penelitian ini, RQ difokuskan untuk mengidentifikasi karakteristik penelitian serta menganalisis penerapan SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning* dalam optimasi model klasifikasi pada *Imbalanced Data*.

Berdasarkan tujuan penelitian, empat *Research Question* yang digunakan disajikan pada Tabel 1.

Tabel 1. *Research Question* Penelitian

Kode	<i>Research Question</i>	Tujuan Analisis
RQ1	Bagaimana karakteristik penelitian mengenai optimasi model klasifikasi pada <i>Imbalanced Data</i> menggunakan SMOTE, <i>Hyperparameter Tuning</i> , dan <i>Ensemble Learning</i> ?	Mengidentifikasi perkembangan penelitian berdasarkan tahun publikasi, domain aplikasi, jurnal, dan metode yang digunakan.
RQ2	Bagaimana penerapan <i>Synthetic Minority Oversampling Technique</i> (SMOTE) dalam meningkatkan performa model klasifikasi pada <i>Imbalanced Data</i> ?	Menganalisis penggunaan SMOTE serta pengaruhnya terhadap penanganan <i>Imbalanced Data</i> dan performa model.
RQ3	Bagaimana metode <i>Hyperparameter Tuning</i> digunakan untuk mengoptimalkan performa model klasifikasi?	Mengidentifikasi metode optimasi parameter serta kecenderungan penerapannya pada berbagai penelitian.
RQ4	Bagaimana penerapan <i>Ensemble Learning</i> dalam meningkatkan performa model klasifikasi pada <i>Imbalanced Data</i> ?	Menganalisis pendekatan <i>Ensemble Learning</i> serta kontribusinya terhadap peningkatan performa model.

Berdasarkan Tabel 1, setiap *Research Question* merepresentasikan aspek analisis yang berbeda namun saling melengkapi. RQ1 digunakan untuk mengidentifikasi karakteristik penelitian, sedangkan RQ2 hingga RQ4 berfokus pada analisis penerapan SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning*. Seluruh RQ tersebut menjadi acuan dalam proses sintesis hasil sehingga setiap analisis dilakukan secara sistematis sesuai dengan tujuan penelitian.

2.4. Strategi Pencarian Literatur

Strategi pencarian literatur disusun untuk memperoleh artikel yang relevan dengan tujuan penelitian sekaligus meminimalkan potensi bias selama proses seleksi [22]. Penyusunan strategi pencarian didasarkan pada *Research Question* yang telah dirumuskan sehingga seluruh artikel yang diperoleh memiliki keterkaitan langsung dengan topik optimasi model klasifikasi pada *Imbalanced Data*.

Penelitian ini menggunakan Scopus sebagai sumber utama literatur karena menyediakan cakupan jurnal internasional bereputasi yang luas serta telah melalui proses peer review. Proses pencarian dilakukan menggunakan fasilitas Advanced Search dengan memanfaatkan kolom TITLE-ABS-KEY, sehingga pencarian difokuskan pada judul, abstrak, dan kata kunci artikel untuk meningkatkan relevansi hasil pencarian.

Konsep utama yang digunakan dalam penyusunan kata kunci meliputi *Imbalanced Data*, SMOTE, *Hyperparameter Tuning*, *Ensemble Learning*, dan classification. Setiap konsep direpresentasikan menggunakan beberapa kata kunci yang dihubungkan dengan operator OR, sedangkan hubungan antar konsep menggunakan operator AND. Rincian konsep dan kata kunci yang digunakan ditunjukkan pada Tabel 2.

Tabel 2. Konsep dan Kata Kunci Pencarian Literatur

Konsep	Kata Kunci (<i>Keywords</i>)
<i>Imbalanced Data</i>	" <i>Imbalanced Data</i> ", "class imbalance", "imbalanced classification", "minority class"
Data Balancing	SMOTE, " <i>Synthetic Minority Oversampling Technique</i> "
<i>Hyperparameter Optimization</i>	" <i>Hyperparameter Tuning</i> ", "parameter optimization", " <i>Grid Search</i> ", " <i>Random Search</i> ", " <i>Bayesian Optimization</i> ", Optuna
<i>Ensemble Learning</i>	" <i>Ensemble Learning</i> ", <i>Random Forest</i> , <i>Stacking</i> , <i>Boosting</i> , <i>Bagging</i> , "voting classifier"
Classification Task	classification

Berdasarkan Tabel 2, kombinasi kata kunci selanjutnya diterapkan dalam Boolean Search menggunakan fasilitas Advanced Search pada Scopus. Pendekatan ini bertujuan memastikan artikel yang diperoleh sesuai dengan fokus penelitian. Format Boolean Search yang digunakan disajikan pada Tabel 3.

Tabel 3. Boolean Search pada Scopus

Search Field	Boolean Search
TITLE-ABS-KEY	TITLE-ABS-KEY ((<i>"Imbalanced Data"</i> OR "class imbalance" OR "imbalanced classification" OR "minority class") AND (SMOTE OR " <i>Synthetic Minority Oversampling Technique</i> ") AND (" <i>Hyperparameter Tuning</i> " OR "parameter optimization" OR " <i>Grid Search</i> " OR " <i>Random Search</i> " OR " <i>Bayesian Optimization</i> " OR Optuna) AND (" <i>Ensemble Learning</i> " OR " <i>Random Forest</i> " OR <i>Stacking</i> OR <i>Boosting</i> OR <i>Bagging</i> OR "voting classifier") AND (classification))

Selain menggunakan Boolean Search, proses pencarian juga dibatasi berdasarkan beberapa parameter untuk memastikan artikel yang diperoleh sesuai dengan ruang lingkup penelitian. Parameter tersebut meliputi basis data, metode pencarian, rentang tahun publikasi, jenis dokumen, dan bahasa publikasi sebagaimana ditunjukkan pada Tabel 4.

Tabel 4. Parameter Pencarian Literatur

Parameter	Keterangan
Database	Scopus
Metode Pencarian	Advanced Search
Search Field	TITLE-ABS-KEY
Rentang Publikasi	2022–2026
Jenis Dokumen	Article
Bahasa	English

Parameter pada Tabel 4 digunakan untuk membatasi hasil pencarian agar artikel yang diperoleh sesuai dengan ruang lingkup penelitian. Selanjutnya, seluruh artikel diseleksi menggunakan kriteria inklusi dan eksklusi, kemudian dievaluasi melalui *Quality Assessment* (QA) sesuai pedoman PRISMA 2020 [22]. Tahapan tersebut memastikan bahwa proses identifikasi

hingga penetapan *Primary Studies* dilakukan secara sistematis, transparan, dan dapat direplikasi.

2.5. Kriteria Inklusi dan Eksklusi

Kriteria inklusi dan eksklusi digunakan untuk memastikan bahwa artikel yang dianalisis memiliki relevansi dengan tujuan penelitian serta memenuhi standar kualitas yang telah ditetapkan. Penerapan kriteria ini bertujuan mengurangi potensi bias selama proses seleksi sehingga hanya artikel yang sesuai dengan ruang lingkup penelitian yang digunakan sebagai *Primary Studies* [22].

Kriteria seleksi yang diterapkan pada penelitian ini disajikan pada Tabel 5.

Tabel 5. Kriteria Inklusi dan Eksklusi

Kriteria	Inklusi	Eksklusi
Topik	Membahas optimasi model klasifikasi pada <i>Imbalanced Data</i> menggunakan SMOTE, <i>Hyperparameter Tuning</i> , atau <i>Ensemble Learning</i> .	Tidak berkaitan dengan optimasi model klasifikasi pada <i>Imbalanced Data</i> .
Tahun Publikasi	2022–2026	Sebelum tahun 2022
Jenis Dokumen	Artikel jurnal (<i>Article</i>)	Prosiding, buku, bab buku, <i>review paper</i> , editorial, dan dokumen selain artikel jurnal
Bahasa	Bahasa Inggris	Selain Bahasa Inggris
Basis Data	Scopus	Basis data selain Scopus
Ketersediaan Dokumen	Teks lengkap (<i>Full Text</i>) tersedia	Teks lengkap tidak tersedia

Berdasarkan Tabel 5, hanya artikel yang memenuhi seluruh kriteria inklusi dan tidak termasuk dalam kriteria eksklusi yang dipertahankan untuk tahap berikutnya. Artikel yang lolos seleksi kemudian dievaluasi melalui *Quality Assessment* (QA) untuk memastikan kualitas metodologi sebelum dilakukan proses ekstraksi data.

2.6. Quality Assessment (QA)

Quality Assessment (QA) dilakukan untuk menilai kualitas metodologi setiap artikel yang telah lolos tahap seleksi. Proses ini bertujuan memastikan bahwa artikel yang digunakan sebagai *Primary Studies* memiliki metodologi yang jelas, relevan dengan tujuan penelitian, serta menyediakan informasi yang memadai untuk proses sintesis [22].

Penilaian kualitas dilakukan menggunakan sejumlah pertanyaan evaluasi yang disusun berdasarkan tujuan penelitian. Setiap artikel dinilai menggunakan kategori "Ya", "Sebagian", atau "Tidak" sebagai dasar dalam menentukan kelayakan artikel untuk dianalisis lebih lanjut.

Tabel 6. Kriteria *Quality Assessment*

Kode	Pertanyaan <i>Quality Assessment</i>
QA1	Apakah penelitian membahas optimasi model klasifikasi pada <i>Imbalanced Data</i> ?
QA2	Apakah penelitian menerapkan SMOTE, <i>Hyperparameter Tuning</i> , atau <i>Ensemble Learning</i> ?
QA3	Apakah metodologi penelitian dijelaskan secara jelas?
QA4	Apakah hasil evaluasi model disajikan secara lengkap?
QA5	Apakah penelitian memiliki keterkaitan dengan <i>Research Question</i> yang telah ditetapkan?

Berdasarkan Tabel 6, artikel dinyatakan layak apabila memenuhi sebagian besar kriteria penilaian. Proses *Quality Assessment* memastikan bahwa hanya artikel dengan metodologi yang memadai dan relevan dengan fokus penelitian yang digunakan pada tahap ekstraksi data dan sintesis hasil.

2.7. Teknik Ekstraksi dan Analisis Data

Proses ekstraksi data dilakukan terhadap seluruh artikel yang telah memenuhi kriteria inklusi, eksklusi, dan *Quality Assessment* (QA). Tahap ini bertujuan mengumpulkan informasi yang relevan untuk menjawab setiap *Research Question* (RQ) secara sistematis. Informasi yang diekstraksi mencakup karakteristik penelitian, metode optimasi yang digunakan, serta hasil evaluasi model sehingga memudahkan proses sintesis dan analisis. Variabel yang diekstraksi dari setiap artikel disajikan pada Tabel 7.

Tabel 7. Variabel Ekstraksi Data

Variabel	Deskripsi
Identitas Artikel	Penulis, tahun publikasi, dan nama jurnal
Domain Penelitian	Bidang atau objek penelitian yang digunakan
Algoritma Klasifikasi	Algoritma yang diterapkan pada penelitian
Teknik Penanganan <i>Imbalanced Data</i>	Metode penyeimbangan data, seperti SMOTE atau variasinya
Metode <i>Hyperparameter Tuning</i>	Pendekatan optimasi parameter yang digunakan
Pendekatan <i>Ensemble Learning</i>	Jenis <i>Ensemble Learning</i> yang diterapkan
Metrik Evaluasi	<i>Accuracy</i> , <i>Precision</i> , <i>Recall</i> , <i>F1-Score</i> , <i>AUC</i> , atau metrik lainnya
Hasil Penelitian	Temuan utama dan performa model yang dilaporkan

Berdasarkan Tabel 7, data yang telah diekstraksi dianalisis menggunakan pendekatan deskriptif dan komparatif. Analisis deskriptif digunakan untuk mengidentifikasi karakteristik penelitian, sedangkan analisis komparatif digunakan untuk membandingkan penerapan SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning* pada berbagai domain penelitian. Hasil analisis tersebut kemudian disintesis untuk menjawab setiap *Research Question* serta mengidentifikasi kecenderungan metode optimasi yang digunakan pada model klasifikasi *Imbalanced Data*.

3. HASIL

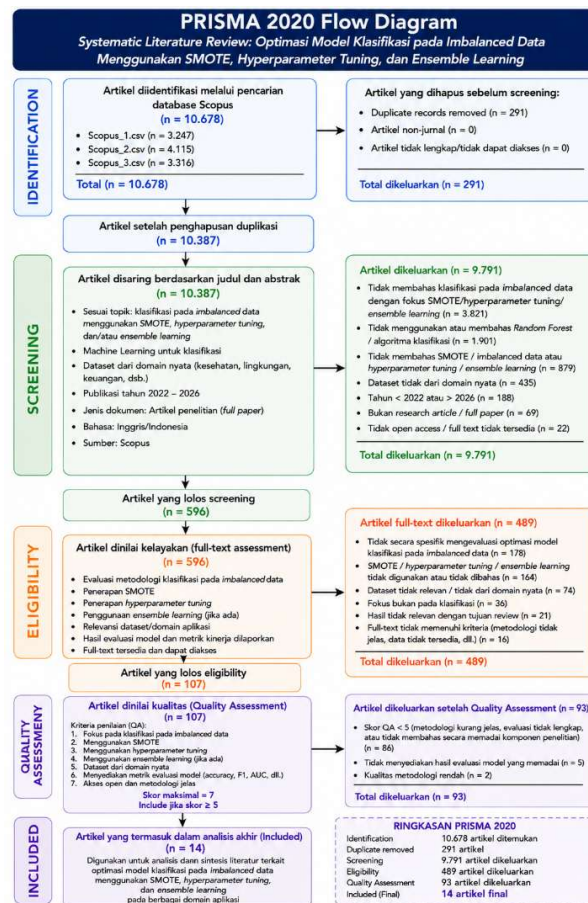
3.1. Hasil Seleksi Literatur

Seleksi literatur dilakukan secara sistematis mengacu pada pedoman PRISMA 2020 untuk memperoleh artikel yang relevan dengan tujuan penelitian [22]. Proses seleksi meliputi empat tahapan, yaitu identifikasi, screening, eligibility, dan penetapan *Primary Studies*. Pendekatan ini memastikan bahwa artikel yang dianalisis memenuhi kriteria inklusi, eksklusi, serta *Quality Assessment* (QA) yang telah ditetapkan pada Bab 2.

Proses identifikasi diawali dengan pencarian artikel pada basis data Scopus menggunakan strategi pencarian yang telah dirancang berdasarkan *Research Question*. Artikel hasil pencarian kemudian disaring melalui pemeriksaan judul, abstrak, dan informasi bibliografi untuk menghilangkan artikel yang tidak relevan maupun duplikat. Artikel yang lolos selanjutnya dievaluasi pada tahap eligibility berdasarkan isi artikel, kriteria seleksi, dan hasil *Quality Assessment*.

Berdasarkan keseluruhan proses seleksi, diperoleh 14 artikel yang memenuhi seluruh kriteria dan ditetapkan sebagai *Primary Studies*. Artikel tersebut berasal dari berbagai domain, seperti kesehatan, kualitas air, deteksi penipuan, rekayasa perangkat lunak, bioinformatika, geoteknik, dan klasifikasi kompetensi kerja. Seluruh artikel digunakan sebagai dasar dalam proses ekstraksi data dan sintesis hasil penelitian.

Alur seleksi literatur pada penelitian ini disajikan pada Gambar 2.



Gambar 2. Diagram PRISMA 2020 Proses Seleksi Literatur

Gambar 2 menunjukkan proses seleksi literatur mulai dari identifikasi artikel pada Scopus hingga penetapan *Primary Studies*. Setiap artikel melewati tahap identifikasi, screening, eligibility, dan *Quality Assessment* sehingga artikel yang digunakan dalam sintesis telah memenuhi kriteria penelitian sesuai pedoman PRISMA 2020 [22].

3.2. Karakteristik Artikel Terpilih

Karakteristik artikel dianalisis untuk mengidentifikasi perkembangan penelitian mengenai optimasi model klasifikasi pada *Imbalanced Data* sekaligus menjawab RQ1. Analisis dilakukan terhadap 14 artikel yang telah memenuhi seluruh tahapan seleksi sehingga memberikan gambaran mengenai tren publikasi, domain penelitian, serta metode optimasi yang digunakan. Karakteristik artikel terpilih disajikan pada Tabel 8.

Berdasarkan Tabel 8, sebagian besar artikel dipublikasikan pada periode 2025–2026, dengan dominasi publikasi pada tahun 2026. Hal ini menunjukkan bahwa optimasi model klasifikasi pada *Imbalanced Data* masih menjadi topik penelitian yang berkembang. Dari sisi domain, bidang kesehatan menjadi aplikasi yang paling banyak dikaji, diikuti kualitas air, deteksi penipuan, rekayasa perangkat lunak, bioinformatika, geoteknik, dan klasifikasi kompetensi kerja.

Tabel 8. Karakteristik Artikel Terpilih

N o	Penulis	Tah un	Nama Jurnal	Domain Penelitian	Metode Utama
1	Mutiana Pratiwi et al.	2026	Journal of Applied Data Sciences	Klasifikasi Kompetensi Kerja	Hyperparameter Tuning, Ensemble Learning
2	Parvez Rahi et al.	2025	IEEE Access	Kesehatan	Random Forest, SMOTE
3	Mahesh Kumbhar et al.	2026	Springer	Optimasi Imbalanced Data	SMOTE, Hyperparameter Tuning
4	Nazim Uddin et al.	2026	Discover Computing	Kesehatan	SMOTE, Hyperparameter Tuning, Ensemble Learning
5	Dadang Priyanto et al.	2025	Engineering, Technology & Applied Science Research	Kesehatan	Random Forest, SMOTE
6	Maram Alamri & Mourad Ykhlef	2026	PeerJ Computer Science	Deteksi Penipuan	Hyperparameter Tuning, Ensemble Learning
7	Jahanur Biswas et al.	2026	Discover Artificial Intelligence	Kesehatan	SMOTE, Hyperparameter Tuning

8	Rasel Ahmed et al.	2026	Discover Artificial Intelligence	Optimasi Model Klasifikasi	Hyperparameter Tuning, Ensemble Learning
9	S. Karthik et al.	2026	Discover Artificial Intelligence	Rekayasa Perangkat Lunak	SMOTE, Hyperparameter Tuning
10	Faris Hassan et al.	2026	Intelligent Medicine	Bioinformatica	Hyperparameter Tuning, Ensemble Learning
11	Pratikumar Parmar et al.	2026	SSRG International Journal of Electronics and Communication Engineering	Kesehatan	Hyperparameter Tuning, Ensemble Learning
12	Rahmi Fadhillah et al.	2025	Engineering, Technology & Applied Science Research	Kualitas Air	Random Forest, SMOTE
13	Qinzhen g Wu et al.	2025	Applied Sciences	Geoteknik	Random Forest, SMOTE, Hyperparameter Tuning
14	Rohit Srivastava et al.	2026	Engineering, Technology & Applied Science Research	Kualitas Air	Hyperparameter Tuning, Ensemble Learning

Sebagian besar penelitian mengombinasikan lebih dari satu teknik optimasi. SMOTE digunakan untuk menangani ketidakseimbangan data, *Hyperparameter Tuning* untuk memperoleh konfigurasi model yang optimal, sedangkan *Ensemble Learning* dimanfaatkan untuk meningkatkan akurasi dan kemampuan generalisasi model. Temuan ini menunjukkan kecenderungan penelitian menuju integrasi beberapa teknik optimasi dibandingkan penggunaan satu metode secara terpisah.

3.3. Analisis Penerapan SMOTE pada Imbalanced Data

Penerapan teknik penanganan *Imbalanced Data* menjadi salah satu strategi utama dalam meningkatkan performa model klasifikasi. Dari 14 artikel yang dianalisis, sebanyak delapan penelitian secara eksplisit menerapkan metode *Oversampling*, dengan SMOTE sebagai teknik yang paling dominan. Selain SMOTE konvensional, beberapa penelitian mengembangkan variasi metode, seperti K-Means SMOTE-ENN, k-Medoids SMOTE, dan pendekatan *Oversampling* berbasis optimasi untuk menghasilkan distribusi data yang lebih representatif.

Ringkasan penerapan teknik penanganan *Imbalanced Data* pada artikel terpilih disajikan pada Tabel 9.

Tabel 9. Penerapan Teknik Penanganan *Imbalanced Data*

No	Penulis	Teknik Penanganan <i>Imbalanced Data</i>	Kombinasi Metode	Dampak terhadap Performa
1	Parvez Rahi et al.	SMOTE	Random Forest + Polynomial Feature Engineering	Meningkatkan performa deteksi gangguan gastrointestinal.
2	Mahesh Kumbhar et al.	Subspace Oversampling	Bayesian Reinforcement	Mengoptimalkan distribusi data minoritas.
3	Nazim Uddin et al.	SMOTE	Ensemble Learning	Meningkatkan klasifikasi penyakit kardiovaskular.
4	Dadang Priyanto et al.	K-Means SMOTE-ENN	Random Forest + PCA	Meningkatkan akurasi klasifikasi data kesehatan.
5	Jahanur Biswas et al.	SMOTE	Randomized Hyperparameter Tuning	Meningkatkan deteksi Alzheimer.
6	S. Karthik et al.	SMOTE	Feature Selection	Meningkatkan prediksi software fault.
7	Rahmi Fadhillah et al.	SMOTE	Resampling Technique	Meningkatkan prediksi kualitas air.
8	Qinzhen g Wu et al.	k-Medoids SMOTE	Random Forest	Meningkatkan klasifikasi rockburst.

Berdasarkan Tabel 9, SMOTE merupakan teknik yang paling banyak digunakan untuk menangani ketidakseimbangan distribusi kelas. Penerapannya umumnya dikombinasikan dengan algoritma klasifikasi serta teknik optimasi lain, seperti *Hyperparameter Tuning*, *Feature Selection*, dan *Ensemble Learning*. Temuan ini menunjukkan bahwa penanganan *Imbalanced Data* telah menjadi bagian dari strategi optimasi model secara menyeluruh.

Selain SMOTE konvensional, beberapa penelitian mengembangkan variasi metode, seperti K-Means SMOTE-ENN dan k-Medoids SMOTE, untuk menghasilkan distribusi data yang lebih representatif. Pendekatan tersebut meningkatkan kemampuan model dalam mengenali kelas minoritas tanpa mengurangi performa pada kelas mayoritas. Hasil sintesis menunjukkan bahwa integrasi SMOTE dengan teknik optimasi lainnya memberikan kontribusi positif terhadap peningkatan performa model klasifikasi pada berbagai domain penelitian.

3.4. Analisis Hyperparameter Tuning

Hyperparameter Tuning merupakan salah satu strategi yang banyak diterapkan untuk mengoptimalkan performa model klasifikasi pada *Imbalanced Data*. Berdasarkan hasil sintesis terhadap 14 artikel, sebanyak 11 penelitian menerapkan optimasi parameter menggunakan berbagai pendekatan, seperti *Optuna*, *Bayesian Optimization*, *Randomized Hyperparameter Tuning*, dan algoritma berbasis metaheuristic. Perkembangan ini menunjukkan bahwa proses optimasi parameter telah beralih dari

penentuan secara manual menuju metode yang lebih adaptif dan otomatis.

Ringkasan penerapan *Hyperparameter Tuning* pada artikel terpilih disajikan pada Tabel 10.

Tabel 10. Penerapan *Hyperparameter Tuning* pada Artikel Terpilih

N o	Penulis	Metode <i>Hyperparameter Tuning</i>	Model yang Dioptimasi	Tujuan Optimasi
1	Mutiana Pratiwi <i>et al.</i>	Optuna	<i>Stacking Ensemble</i>	Memperoleh kombinasi parameter optimal untuk meningkatkan akurasi klasifikasi.
2	Mahesh Kumbhar <i>et al.</i>	<i>Bayesian Optimization</i>	<i>Framework Oversampling</i>	Mengoptimalkan proses pembangkitan data dan performa klasifikasi.
3	Nazim Uddin <i>et al.</i>	<i>Hyperparameter Optimization</i>	<i>Ensemble Learning</i>	Meningkatkan performa prediksi penyakit kardiovaskular.
4	Maram Alamri & Mourad Ykhlef	Optimized Classification	<i>Hybrid Machine Learning</i>	Mengoptimalkan konfigurasi model untuk deteksi penipuan.
5	Jahanur Biswas <i>et al.</i>	<i>Randomized Hyperparameter Tuning</i>	<i>Machine Learning Classifier</i>	Meningkatkan performa deteksi Alzheimer.
6	Rasel Ahmed <i>et al.</i>	<i>Hyperparameter Optimization</i>	<i>Soft Voting Ensemble</i>	Mengoptimalkan bobot model <i>ensemble</i> .
7	S. Karthik <i>et al.</i>	<i>ABC-Based Optimization</i>	Software Fault Prediction Model	Mengoptimalkan proses klasifikasi <i>software fault</i> .
8	Faris Hassan <i>et al.</i>	<i>Metaheuristic Optimization</i>	<i>Stacked Ensemble</i>	Meningkatkan performa prediksi resistensi inhibitor ABL1.
9	Pratikumar Parmar <i>et al.</i>	Optimized <i>Ensemble Classifier</i>	<i>Ensemble Classifier</i>	Meningkatkan performa klasifikasi penyakit jantung.
10	Qinzheng Wu <i>et al.</i>	<i>BSLO Optimization</i>	<i>Random Forest</i>	Mengoptimalkan parameter <i>Random Forest</i> .
11	Rohit Srivastava <i>et al.</i>	<i>Hyperparameter Optimization</i>	<i>Ensemble Machine Learning</i>	Meningkatkan performa prediksi kualitas air.

Berdasarkan Tabel 10, *Hyperparameter Tuning* diterapkan pada berbagai algoritma klasifikasi maupun model ensemble untuk memperoleh konfigurasi parameter yang sesuai dengan karakteristik data. Selain meningkatkan nilai *Accuracy*, beberapa penelitian juga melaporkan peningkatan pada *Precision*, *Recall*, dan *F1-Score*, terutama ketika optimasi parameter dikombinasikan dengan teknik penanganan *Imbalanced Data*.

Hasil sintesis menunjukkan bahwa pendekatan optimasi yang lebih adaptif, seperti *Optuna*, *Bayesian Optimization*, dan algoritma berbasis metaheuristic, semakin banyak digunakan dibandingkan proses penentuan parameter secara manual. Pendekatan tersebut memungkinkan proses eksplorasi ruang pencarian parameter dilakukan secara lebih efisien sehingga berpotensi menghasilkan konfigurasi model yang lebih optimal. Temuan ini menunjukkan bahwa *Hyperparameter Tuning* telah menjadi komponen penting dalam optimasi model klasifikasi, meskipun efektivitas setiap metode tetap dipengaruhi oleh karakteristik dataset dan algoritma yang digunakan.

3.5. Analisis Ensemble Learning

Ensemble Learning merupakan salah satu pendekatan yang banyak digunakan untuk meningkatkan performa model klasifikasi pada *Imbalanced Data*. Berdasarkan hasil sintesis terhadap 14 artikel, sebanyak tujuh penelitian menerapkan berbagai pendekatan ensemble, seperti *Stacking Ensemble*, *Soft Voting Ensemble*, *Hybrid Ensemble*, dan *Ensemble Classifier*. Temuan ini menunjukkan bahwa penelitian terkini semakin mengintegrasikan beberapa model klasifikasi untuk memperoleh prediksi yang lebih akurat dan memiliki kemampuan generalisasi yang lebih baik.

Ringkasan penerapan *Ensemble Learning* pada artikel terpilih disajikan pada Tabel 11.

Tabel 11. Penerapan *Ensemble Learning* pada Artikel Terpilih

N o	Penulis	Pendekatan <i>Ensemble Learning</i>	Kombinasi Metode	Tujuan Penerapan
1	Mutiana Pratiwi <i>et al.</i>	<i>Stacking Ensemble</i>	Optuna	Meningkatkan akurasi klasifikasi kompetensi kerja melalui optimasi model <i>ensemble</i> .
2	Nazim Uddin <i>et al.</i>	<i>Explainable Ensemble Learning</i>	SMOTE + Feature Optimization	Meningkatkan performa prediksi penyakit kardiovaskular.
3	Maram Alamri & Mourad Ykhlef	<i>Hybrid Ensemble</i>	<i>Hyperparameter Optimization</i>	Mengoptimalkan deteksi penipuan transaksi.
4	Rasel Ahmed <i>et al.</i>	<i>Soft Voting Ensemble</i>	<i>Hyperparameter Optimization</i>	Mengoptimalkan bobot model <i>ensemble</i> untuk meningkatkan performa klasifikasi.

5	Faris Hassan et al.	Stacked Ensemble	Metaheuristic Optimization	Meningkatkan akurasi prediksi resistensi inhibitor ABL1.
6	Pratikkumar Parmar et al.	Ensemble Classifier	Hyperparameter Optimization	Meningkatkan performa klasifikasi penyakit jantung.
7	Rohit Srivastava et al.	Ensemble Machine Learning	Feature Engineering + Hyperparameter Optimization	Meningkatkan performa prediksi kualitas air.

Berdasarkan Tabel 11, *Ensemble Learning* umumnya diterapkan bersama teknik optimasi lain, seperti *Hyperparameter Tuning*, *SMOTE*, dan *Feature Selection*. Integrasi tersebut menunjukkan bahwa peningkatan performa model dipengaruhi oleh kombinasi metode ensemble, kualitas data, dan proses optimasi yang diterapkan pada tahap sebelumnya. Selain meningkatkan nilai *Accuracy*, beberapa penelitian juga melaporkan peningkatan pada *Precision*, *Recall*, dan *F1-Score*.

Hasil sintesis menunjukkan bahwa pendekatan *Stacking Ensemble* menjadi metode yang paling banyak diterapkan, diikuti *Soft Voting Ensemble* dan berbagai model *Hybrid Ensemble*. Pendekatan tersebut mengombinasikan keunggulan beberapa algoritma klasifikasi sehingga menghasilkan model dengan kemampuan generalisasi yang lebih baik. Temuan ini

menunjukkan bahwa *Ensemble Learning* telah menjadi salah satu komponen penting dalam optimasi model klasifikasi, terutama ketika diintegrasikan dengan teknik penanganan *Imbalanced Data* dan *Hyperparameter Tuning*.

3.6. Sintesis Hasil Penelitian dan Research Gap

Hasil sintesis terhadap 14 artikel menunjukkan bahwa penelitian mengenai optimasi model klasifikasi pada *Imbalanced Data* berkembang menuju integrasi berbagai teknik optimasi. Sebagian besar penelitian tidak lagi mengandalkan satu pendekatan, tetapi mengombinasikan *SMOTE*, *Hyperparameter Tuning*, dan *Ensemble Learning* untuk meningkatkan kemampuan model dalam mengenali kelas minoritas sekaligus mempertahankan performa klasifikasi secara keseluruhan. Temuan ini menunjukkan adanya pergeseran dari optimasi berbasis algoritma tunggal menuju strategi yang lebih terintegrasi.

Selain menunjukkan perkembangan metode, hasil sintesis juga mengidentifikasi beberapa peluang penelitian yang masih terbuka. Variasi teknik penanganan *Imbalanced Data*, metode *Hyperparameter Tuning*, serta pendekatan *Ensemble Learning* menunjukkan bahwa belum terdapat strategi optimasi yang secara konsisten memberikan performa terbaik pada seluruh karakteristik dataset. Di sisi lain, sebagian besar penelitian masih berfokus pada peningkatan metrik performa, sementara aspek interpretabilitas model, efisiensi komputasi, dan implementasi pada lingkungan nyata masih relatif terbatas.

Ringkasan hasil sintesis dan identifikasi *Research Gap* disajikan pada Tabel 12.

Tabel 12. Sintesis Hasil Penelitian dan *Research Gap*

Aspek	Hasil Sintesis	Research Gap
Penanganan <i>Imbalanced Data</i>	SMOTE menjadi teknik yang paling banyak digunakan dan menunjukkan peningkatan performa klasifikasi.	Masih terbatas penelitian yang membandingkan efektivitas berbagai variasi SMOTE pada karakteristik dataset yang berbeda.
<i>Hyperparameter Tuning</i>	Berbagai pendekatan optimasi digunakan untuk meningkatkan performa model.	Belum terdapat metode optimasi yang secara konsisten memberikan performa terbaik pada seluruh karakteristik dataset.
<i>Ensemble Learning</i>	Pendekatan <i>ensemble</i> banyak dikombinasikan dengan teknik optimasi lainnya.	Integrasi <i>Ensemble Learning</i> , <i>SMOTE</i> , dan <i>Hyperparameter Tuning</i> secara bersamaan masih relatif terbatas.
Domain Penelitian	Penelitian didominasi bidang kesehatan, kualitas air, dan deteksi penipuan.	Diperlukan validasi pada domain aplikasi lain untuk menguji kemampuan generalisasi metode yang diusulkan.
Evaluasi Model	Sebagian besar penelitian menggunakan <i>Accuracy</i> , <i>Precision</i> , <i>Recall</i> , dan <i>F1-Score</i> .	Evaluasi terhadap interpretabilitas model, efisiensi komputasi, dan implementasi pada lingkungan nyata masih terbatas.

Berdasarkan Tabel 12, sebagian besar penelitian berorientasi pada peningkatan performa model melalui integrasi beberapa teknik optimasi. Namun, hasil sintesis menunjukkan bahwa efektivitas setiap pendekatan masih dipengaruhi oleh karakteristik dataset, algoritma klasifikasi, dan domain aplikasi yang digunakan. Kondisi tersebut mengindikasikan bahwa belum terdapat pendekatan optimasi yang dapat diterapkan secara universal pada seluruh permasalahan *Imbalanced Data*.

Secara keseluruhan, hasil sintesis memberikan gambaran mengenai perkembangan penelitian sekaligus mengidentifikasi peluang pengembangan pada aspek integrasi metode,

interpretabilitas model, efisiensi komputasi, dan perluasan domain aplikasi. Temuan tersebut menjadi dasar dalam pembahasan mengenai karakteristik penelitian serta efektivitas penerapan *SMOTE*, *Hyperparameter Tuning*, dan *Ensemble Learning* pada bagian selanjutnya.

4. PEMBAHASAN

4.1. Pembahasan RQ1: Karakteristik Penelitian

Hasil sintesis menunjukkan bahwa penelitian mengenai optimasi model klasifikasi pada *Imbalanced Data* mengalami perkembangan yang signifikan pada periode 2025–2026. Meskipun proses pencarian literatur mencakup publikasi tahun 2022–2026, seluruh artikel yang memenuhi kriteria seleksi berasal dari dua tahun terakhir. Kondisi ini mengindikasikan bahwa penerapan SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning* masih menjadi fokus utama dalam pengembangan model klasifikasi untuk menangani distribusi data yang tidak seimbang [5], [6].

Selain mengalami peningkatan dari sisi publikasi, penelitian juga menunjukkan keberagaman domain aplikasi, meliputi kesehatan, deteksi penipuan, kualitas air, rekayasa perangkat lunak, geoteknik, hingga klasifikasi kompetensi kerja [10], [16]. Temuan tersebut memperlihatkan bahwa permasalahan *Imbalanced Data* tidak terbatas pada satu bidang tertentu, tetapi muncul pada berbagai jenis data yang memerlukan model klasifikasi dengan kemampuan prediksi yang andal.

Perkembangan lain yang terlihat adalah perubahan strategi optimasi model. Penelitian terkini tidak lagi berfokus pada penggunaan satu algoritma klasifikasi, melainkan mengintegrasikan beberapa teknik optimasi secara bersamaan [13], [21]. Kombinasi SMOTE sebagai metode penyeimbangan data, *Hyperparameter Tuning* untuk memperoleh konfigurasi model yang optimal, serta *Ensemble Learning* sebagai pendekatan peningkatan performa menunjukkan kecenderungan menuju proses optimasi yang lebih komprehensif. Pendekatan tersebut tidak hanya meningkatkan *Accuracy*, tetapi juga memperbaiki kemampuan model dalam mengenali kelas minoritas serta meningkatkan kemampuan generalisasi.

Hasil sintesis juga menunjukkan bahwa pemilihan teknik optimasi dipengaruhi oleh karakteristik dataset. Dataset dengan tingkat ketidakseimbangan yang tinggi umumnya memerlukan proses *Oversampling*, sedangkan dataset dengan karakteristik fitur yang lebih kompleks lebih banyak memanfaatkan *Hyperparameter Tuning* maupun *Ensemble Learning* untuk memperoleh performa yang optimal [14]. Temuan ini menunjukkan bahwa belum terdapat satu strategi optimasi yang dapat diterapkan secara universal pada seluruh permasalahan klasifikasi.

Secara keseluruhan, RQ1 menunjukkan bahwa penelitian mengenai optimasi model klasifikasi pada *Imbalanced Data* berkembang menuju integrasi berbagai teknik optimasi. SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning* tidak lagi diterapkan secara terpisah, tetapi saling melengkapi untuk meningkatkan akurasi, kemampuan generalisasi, dan performa model pada berbagai domain aplikasi.

4.2. Pembahasan RQ2: Efektivitas Penerapan SMOTE pada Imbalanced Data

Hasil sintesis menunjukkan bahwa *Synthetic Minority Oversampling Technique* (SMOTE) merupakan metode yang

paling banyak digunakan untuk menangani *Imbalanced Data*. Dominasi penggunaan SMOTE menunjukkan bahwa teknik ini masih efektif dalam meningkatkan representasi kelas minoritas sebelum proses pelatihan model dilakukan [10], [11]. Berbagai penelitian melaporkan bahwa penerapan SMOTE mampu meningkatkan kemampuan model dalam mengenali kelas minoritas sehingga menghasilkan performa klasifikasi yang lebih seimbang dibandingkan penggunaan data dengan distribusi kelas yang tidak proporsional [8], [18].

Efektivitas SMOTE tidak hanya dipengaruhi oleh proses *Oversampling*, tetapi juga oleh karakteristik dataset dan metode klasifikasi yang digunakan. Hasil sintesis menunjukkan bahwa SMOTE umumnya dikombinasikan dengan algoritma *Random Forest*, *Hyperparameter Tuning*, maupun *Feature Selection* untuk mengurangi bias terhadap kelas mayoritas dan meningkatkan akurasi prediksi [10], [14]. Temuan ini menunjukkan bahwa efektivitas penanganan *Imbalanced Data* dipengaruhi oleh integrasi antara teknik penyeimbangan data dan strategi optimasi model.

Perkembangan penelitian juga menunjukkan adanya variasi penerapan SMOTE, seperti K-Means SMOTE-ENN dan k-Medoids SMOTE, yang dirancang untuk menghasilkan distribusi data sintetis yang lebih representatif [11], [14]. Pendekatan tersebut tidak hanya berfokus pada penambahan jumlah sampel kelas minoritas, tetapi juga pada peningkatan kualitas data hasil *Oversampling*, sehingga mampu meningkatkan performa model pada berbagai karakteristik dataset.

Selain itu, efektivitas SMOTE umumnya dievaluasi menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Meskipun seluruh metrik menunjukkan peningkatan setelah proses penyeimbangan data, *Recall* dan *F1-Score* menjadi indikator yang lebih relevan untuk mengevaluasi performa model pada kondisi *Imbalanced Data* karena lebih mencerminkan kemampuan model dalam mengidentifikasi kelas minoritas [6], [18].

Secara keseluruhan, RQ2 menunjukkan bahwa SMOTE masih menjadi pendekatan yang efektif dalam menangani *Imbalanced Data*. Namun, peningkatan performa yang lebih optimal diperoleh ketika SMOTE diintegrasikan dengan *Hyperparameter Tuning*, *Ensemble Learning*, atau teknik optimasi lainnya. Hal ini menunjukkan bahwa keberhasilan model klasifikasi dipengaruhi oleh kombinasi metode yang digunakan serta karakteristik dataset yang dianalisis.

4.3. Pembahasan RQ3: Peran Hyperparameter Tuning dalam Optimasi Model

Hasil sintesis menunjukkan bahwa *Hyperparameter Tuning* menjadi salah satu komponen penting dalam optimasi model klasifikasi pada *Imbalanced Data*. Sebagian besar penelitian tidak lagi menggunakan konfigurasi parameter bawaan, tetapi menerapkan berbagai pendekatan optimasi untuk memperoleh kombinasi parameter yang lebih sesuai dengan karakteristik dataset [7], [13]. Temuan ini menunjukkan bahwa peningkatan performa model tidak hanya ditentukan oleh pemilihan algoritma klasifikasi, tetapi juga oleh strategi optimasi parameter yang diterapkan sebelum proses pelatihan.

Perkembangan penelitian memperlihatkan meningkatnya penggunaan metode optimasi yang lebih adaptif, seperti *Optuna*, *Bayesian Optimization*, *Randomized Hyperparameter Tuning*, dan algoritma berbasis metaheuristic [7], [13], [19]. Pergeseran tersebut menunjukkan bahwa proses pencarian parameter telah berkembang dari pendekatan manual menuju teknik yang lebih otomatis dan efisien. Dengan demikian, *Hyperparameter Tuning* telah menjadi bagian penting dalam proses optimasi model, bukan lagi sekadar tahap pelengkap.

Selain meningkatkan nilai *Accuracy*, penerapan *Hyperparameter Tuning* juga berdampak pada peningkatan *Precision*, *Recall*, dan *F1-Score* [5], [6]. Peningkatan tersebut menunjukkan bahwa optimasi parameter tidak hanya memperbaiki ketepatan prediksi secara umum, tetapi juga meningkatkan kemampuan model dalam mengenali kelas minoritas. Oleh karena itu, *Hyperparameter Tuning* berperan dalam meningkatkan kemampuan generalisasi model, terutama ketika dikombinasikan dengan teknik penanganan *Imbalanced Data*.

Meskipun demikian, hasil sintesis menunjukkan bahwa belum terdapat satu metode *Hyperparameter Tuning* yang secara konsisten memberikan performa terbaik pada seluruh karakteristik dataset [7], [14]. Efektivitas setiap pendekatan dipengaruhi oleh jenis data, jumlah fitur, distribusi kelas, serta algoritma klasifikasi yang digunakan. Kondisi ini menunjukkan bahwa pemilihan metode optimasi perlu disesuaikan dengan karakteristik permasalahan yang dihadapi.

Secara keseluruhan, RQ3 menunjukkan bahwa *Hyperparameter Tuning* memberikan kontribusi penting dalam meningkatkan performa model klasifikasi pada *Imbalanced Data*. Namun, efektivitasnya tetap dipengaruhi oleh karakteristik dataset dan metode optimasi yang digunakan, sehingga belum terdapat satu pendekatan yang dapat diterapkan secara universal pada seluruh kasus klasifikasi.

4.4. Pembahasan RQ4: Integrasi Ensemble Learning dalam Optimasi Model

Hasil sintesis menunjukkan bahwa *Ensemble Learning* semakin banyak dimanfaatkan sebagai strategi optimasi untuk meningkatkan performa model klasifikasi pada *Imbalanced Data*. Berbeda dengan pendekatan yang hanya mengandalkan satu algoritma, *Ensemble Learning* mengombinasikan beberapa model sehingga mampu menghasilkan prediksi yang lebih stabil dan memiliki kemampuan generalisasi yang lebih baik [13], [21]. Perkembangan ini menunjukkan bahwa penelitian terkini tidak hanya berorientasi pada peningkatan *Accuracy*, tetapi juga pada konsistensi performa model pada berbagai karakteristik dataset.

Sebagian besar penelitian menerapkan *Ensemble Learning* bersama teknik optimasi lain, seperti SMOTE, *Hyperparameter Tuning*, dan *Feature Selection* [5], [6]. Integrasi tersebut menunjukkan bahwa peningkatan performa model dipengaruhi oleh kombinasi metode ensemble, kualitas data, dan proses optimasi yang dilakukan sebelum pelatihan model. Dengan demikian, *Ensemble Learning* berperan sebagai bagian dari strategi optimasi yang lebih komprehensif, bukan sebagai pendekatan yang berdiri sendiri.

Hasil sintesis juga menunjukkan bahwa *Stacking Ensemble* menjadi pendekatan yang paling banyak digunakan, diikuti *Soft Voting Ensemble* dan berbagai model *Hybrid Ensemble* [13], [19]. Pendekatan tersebut mengombinasikan keunggulan beberapa algoritma klasifikasi sehingga mampu meningkatkan kemampuan prediksi dan generalisasi model pada berbagai domain aplikasi.

Meskipun memberikan hasil yang positif, penerapan *Ensemble Learning* masih menghadapi beberapa tantangan. Sebagian besar penelitian berfokus pada peningkatan performa klasifikasi, sedangkan aspek efisiensi komputasi, kompleksitas model, dan interpretabilitas masih relatif terbatas [19], [20]. Kondisi ini menunjukkan bahwa pengembangan metode ensemble pada masa mendatang perlu mempertimbangkan keseimbangan antara peningkatan performa dan kemudahan implementasi pada lingkungan nyata.

Secara keseluruhan, RQ4 menunjukkan bahwa *Ensemble Learning* memberikan kontribusi yang signifikan terhadap optimasi model klasifikasi pada *Imbalanced Data*, terutama ketika diintegrasikan dengan SMOTE dan *Hyperparameter Tuning*. Integrasi berbagai teknik optimasi tersebut mampu meningkatkan akurasi, kemampuan generalisasi, dan stabilitas model, meskipun masih diperlukan penelitian lanjutan untuk menghasilkan pendekatan ensemble yang lebih efisien dan mudah diinterpretasikan.

4.5. Implikasi Penelitian

Hasil sintesis terhadap 14 artikel menunjukkan bahwa optimasi model klasifikasi pada *Imbalanced Data* berkembang dari penggunaan metode tunggal menuju integrasi berbagai teknik optimasi. Kombinasi SMOTE sebagai metode penyeimbangan data, *Hyperparameter Tuning* untuk memperoleh konfigurasi model yang optimal, dan *Ensemble Learning* sebagai strategi peningkatan performa menunjukkan bahwa performa model dipengaruhi tidak hanya oleh algoritma klasifikasi, tetapi juga oleh proses optimasi pada setiap tahapan pengembangannya.

Dari sisi akademik, penelitian ini memberikan gambaran mengenai arah perkembangan penelitian optimasi model klasifikasi pada *Imbalanced Data*. Hasil sintesis menunjukkan bahwa penelitian terkini lebih banyak mengintegrasikan beberapa teknik optimasi dibandingkan mengembangkan satu metode secara terpisah. Selain itu, penelitian ini mengidentifikasi beberapa *Research Gap*, antara lain masih terbatasnya penelitian yang mengombinasikan SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning* dalam satu kerangka optimasi, serta masih minimnya kajian mengenai interpretabilitas model, efisiensi komputasi, dan kemampuan generalisasi pada berbagai domain aplikasi.

Dari sisi praktis, hasil penelitian ini dapat menjadi referensi bagi peneliti maupun praktisi dalam menentukan strategi optimasi yang sesuai dengan karakteristik dataset. Hasil sintesis menunjukkan bahwa belum terdapat satu metode yang dapat diterapkan secara universal pada seluruh kasus *Imbalanced Data*. Oleh karena itu, pemilihan teknik optimasi perlu mempertimbangkan tingkat ketidakseimbangan data,

karakteristik fitur, kompleksitas model, serta tujuan evaluasi yang ingin dicapai [7], [14].

Temuan penelitian juga membuka peluang bagi penelitian selanjutnya untuk mengembangkan kerangka optimasi yang mengintegrasikan penanganan *Imbalanced Data*, *Hyperparameter Tuning*, dan *Ensemble Learning* dalam satu alur pemodelan yang terpadu. Selain peningkatan performa, penelitian mendatang perlu memberikan perhatian lebih pada aspek interpretabilitas model, efisiensi komputasi, serta validasi pada berbagai domain aplikasi agar model yang dihasilkan lebih mudah diimplementasikan pada kondisi nyata [19], [21].

Secara keseluruhan, penelitian ini menunjukkan bahwa optimasi model klasifikasi pada *Imbalanced Data* memerlukan pendekatan yang terintegrasi. Sinergi antara SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning* berpotensi menghasilkan model dengan performa, kemampuan generalisasi, dan keandalan yang lebih baik, sekaligus menjadi dasar bagi pengembangan strategi optimasi yang lebih adaptif pada penelitian berikutnya.

5. KESIMPULAN

Penelitian ini menerapkan *Systematic Literature Review* (SLR) terhadap 14 artikel yang membahas optimasi model klasifikasi pada *Imbalanced Data* melalui penerapan *Synthetic Minority Oversampling Technique* (SMOTE), *Hyperparameter Tuning*, dan *Ensemble Learning*. Hasil sintesis menunjukkan bahwa penelitian terkini cenderung mengintegrasikan ketiga pendekatan tersebut untuk meningkatkan performa model, terutama dalam mengenali kelas minoritas, meningkatkan akurasi, dan memperkuat kemampuan generalisasi.

Berdasarkan hasil kajian, SMOTE masih menjadi teknik yang paling banyak digunakan untuk mengatasi ketidakseimbangan data, sedangkan *Hyperparameter Tuning* berperan dalam memperoleh konfigurasi model yang optimal. Sementara itu, *Ensemble Learning* banyak dimanfaatkan untuk meningkatkan stabilitas dan keandalan prediksi melalui kombinasi beberapa algoritma klasifikasi. Temuan ini menunjukkan bahwa optimasi model tidak lagi bergantung pada satu metode, tetapi pada integrasi berbagai teknik yang saling melengkapi sesuai karakteristik dataset.

Penelitian ini juga mengidentifikasi beberapa *Research Gap*, antara lain masih terbatasnya penelitian yang mengintegrasikan SMOTE, *Hyperparameter Tuning*, dan *Ensemble Learning* dalam satu kerangka optimasi, serta minimnya kajian mengenai interpretabilitas model, efisiensi komputasi, dan validasi pada berbagai domain aplikasi. Kondisi tersebut menunjukkan bahwa pengembangan strategi optimasi yang lebih adaptif dan komprehensif masih menjadi peluang penelitian di masa mendatang.

Secara keseluruhan, penelitian ini memberikan sintesis mengenai perkembangan optimasi model klasifikasi pada *Imbalanced Data* sekaligus menyajikan arah pengembangan penelitian selanjutnya. Penelitian mendatang diharapkan mengembangkan kerangka optimasi yang lebih terintegrasi, memperhatikan aspek interpretabilitas dan efisiensi komputasi, serta memvalidasi

model pada domain aplikasi yang lebih beragam agar menghasilkan model yang lebih andal dan mudah diimplementasikan.

DAFTAR PUSTAKA

- [1] E. Traka, G. Papageorgiou, G. Mantzavinis, and C. Tjortjis, "Comparative Forecasting and Misclassification Analysis Using Health Survey Data," *AI*, vol. 7, no. 4, pp. 1–27, 2026, doi: [10.3390/ai7040148](https://doi.org/10.3390/ai7040148).
- [2] S. A. Salloum, A. Almomani, K. M. Alomari, T. Khmour, and M. Alauthman, "Predicting Thyroid Dysfunction Using Classical *Machine Learning* with Rigorous Statistical Evaluation," *IEEE Access*, vol. 14, no. February, pp. 28852–28866, 2026, doi: [10.1109/ACCESS.2026.3666210](https://doi.org/10.1109/ACCESS.2026.3666210).
- [3] D. W. Brata, A. Djunaidy, D. O. Siahaan, and S. Handoyo, "The Acquiring Optimal Models of *Random Forest* and *Support Vector Machine* Through Tuning Hyperparameters in Classifying the *Imbalanced Data*," *Int. J. Adv. Comput. Sci. Appl.*, vol. 17, no. 2, pp. 739–749, 2026, doi: [10.14569/IJACSA.2026.0170273](https://doi.org/10.14569/IJACSA.2026.0170273).
- [4] A. Arabiat, O. Almomani, and M. Altayeb, "*Machine Learning*-Driven Prediction of EV Charging Behavior for Renewable and Grid Optimization," *J. Robot. Control*, vol. 7, no. 1, pp. 3462–3471, 2026, doi: [10.18196/jrc.v7i1.26225](https://doi.org/10.18196/jrc.v7i1.26225).
- [5] R. Ahmed, N. Fahad, M. S. U. Miah, M. J. Hossen, and K. Bhattacharjee, *HyOPTEnsemble: custom-weighted Soft Voting Hyperparameter Optimization ensemble model, explainable-AI for predicting mental state among university students*, vol. 6, no. 1. 2026. doi: [10.1007/s44163-025-00708-9](https://doi.org/10.1007/s44163-025-00708-9).
- [6] N. Uddin, P. Mahmud, M. Ahammed, M. A. Uddin, M. Kazi, and M. A. Talukder, "Explainable *Ensemble Learning* model for cardiovascular disease prediction with feature optimization and data balancing," *Discov. Comput.*, vol. 29, no. 1, 2026, doi: [10.1007/s10791-026-09930-0](https://doi.org/10.1007/s10791-026-09930-0).
- [7] M. Kumbhar, S. Bandaru, and A. Karlsson, *Imbalanced Data Oversampling through subspace optimization with Bayesian reinforcement*, vol. 59, no. 1. 2026. doi: [10.1007/s10462-025-11417-1](https://doi.org/10.1007/s10462-025-11417-1).
- [8] R. Fadhilah, H. Kuswanto, and D. D. Prastyo, "Performance Evaluation of Classification Methods Utilizing Resampling Techniques for Water Quality Prediction on *Imbalanced Data*," *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 4, pp. 26091–26099, 2025, doi: [10.48084/etasr.11832](https://doi.org/10.48084/etasr.11832).
- [9] S. Seo, D. Oh, K. Kang, H. J. Park, and J. H. Jeon,

- “Construction Accident Prediction via Generative AI and *AutoML* Approaches,” *Appl. Sci.*, vol. 16, no. 5, pp. 1–17, 2026, doi: [10.3390/app16052412](https://doi.org/10.3390/app16052412).
- [10] P. Rahi, M. R. Hussain, S. Alqahtani, M. Husain, A. P. Singh, and I. Singh, “A Novel *Random Forest*-SMOTE Framework With *Polynomial Feature Engineering* for Early Detection of Gastrointestinal Disorders,” *IEEE Access*, vol. 13, no. September, pp. 188577–188604, 2025, doi: [10.1109/ACCESS.2025.3626072](https://doi.org/10.1109/ACCESS.2025.3626072).
- [11] D. Priyanto, H. Hairani, K. Marzuki, and M. Innuddin, “Optimization of *Random Forest* for Health Data Classification Using PCA and K-Means SMOTE-ENN,” *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 5, pp. 27646–27652, 2025, doi: [10.48084/etasr.12976](https://doi.org/10.48084/etasr.12976).
- [12] S. Karthik, S. Kaliraj, and M. D. Lydia, “Cascade framework for software fault prediction using ABC-based *Feature Selection* and SMOTE,” *Discov. Artif. Intell.*, vol. 6, no. 1, 2026, doi: [10.1007/s44163-026-01125-2](https://doi.org/10.1007/s44163-026-01125-2).
- [13] M. Pratiwi, S. Defit, and M. Tajuddin, “Adaptive Integration of Optuna Optimization and *Stacking Ensemble Learning* for Automated Work Competency Classification,” *J. Appl. Data Sci.*, vol. 7, no. 2, pp. 1108–1119, 2026, doi: [10.47738/jads.v7i2.1228](https://doi.org/10.47738/jads.v7i2.1228).
- [14] Q. Wu, B. Dai, D. Li, H. Jia, and P. Li, “Classification of Rockburst Intensity Grades: A Method Integrating k-Medoids-SMOTE and BSLO-RF,” *Appl. Sci.*, vol. 15, no. 16, pp. 1–23, 2025, doi: [10.3390/app15169045](https://doi.org/10.3390/app15169045).
- [15] K. Abnoosian, R. Farnoosh, and M. H. Behzadi, “Prediction of diabetes disease using an ensemble of *Machine Learning* multi-classifier models,” *BMC Bioinformatics*, vol. 24, no. 1, pp. 1–24, 2023, doi: [10.1186/s12859-023-05465-z](https://doi.org/10.1186/s12859-023-05465-z).
- [16] R. Srivastava, K. Sinha, D. Samanta, R. Alsabet, W. El-Shafai, and A. T. Azar, “Ensemble *Machine Learning* for Reliable Water Potability Prediction with Optimized Physicochemical *Feature Engineering*,” *Eng. Technol. Appl. Sci. Res.*, vol. 16, no. 2, pp. 33647–33659, 2026, doi: [10.48084/etasr.16495](https://doi.org/10.48084/etasr.16495).
- [17] T. M. Lam, S. N. Prova, T. e. Mustafa, and N. A. Khan, “Improving Lung Cancer Prediction Using Advanced *Hyperparameter Optimization* and *Nested Cross-Validation*,” *Eng. Reports*, vol. 7, no. 11, pp. 1–10, 2025, doi: [10.1002/eng2.70486](https://doi.org/10.1002/eng2.70486).
- [18] J. Biswas *et al.*, “Performance-optimized Alzheimer’s detection using *Machine Learning* with SMOTE and *Randomized Hyperparameter Tuning*,” *Discov. Artif. Intell.*, 2026, doi: [10.1007/s44163-025-00758-z](https://doi.org/10.1007/s44163-025-00758-z).
- [19] F. Hassan, M. A. Deif, A. Zaghloul, R. Elgohary, and M. Khishe, “Explainable stacked-ensemble prediction of ABL1 tyrosine-kinase inhibitor resistance: A metaheuristic-optimized pipeline,” *Intell. Med.*, vol. 13, no. October 2025, p. 100341, 2026, doi: [10.1016/j.ibmed.2025.100341](https://doi.org/10.1016/j.ibmed.2025.100341).
- [20] P. Parmar, N. Shah, and V. P. Jariwala, “Enhancing Prediction Accuracy for Cardiovascular Heart Diseases through Optimized Ensembled Classifiers,” *SSRG Int. J. Electron. Commun. Eng.*, vol. 13, no. 3, pp. 203–212, 2026, doi: [10.14445/23488549/IJECE-V13I3P117](https://doi.org/10.14445/23488549/IJECE-V13I3P117).
- [21] M. Alamri and M. Ykhlef, “A novel *Hybrid Machine Learning* model for credit card fraud detection using anomaly detection and optimised classification,” *PeerJ Comput. Sci.*, vol. 12, 2026, doi: [10.7717/peerj-cs.3748](https://doi.org/10.7717/peerj-cs.3748).
- [22] M. J. Page *et al.*, “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *Bmj*, vol. 372, 2021, doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71).